Music Symbol Recognition

Joyce Ooi Boon Ee
Faculty of Engineering and Technology
Multimedia University
Melaka, Malaysia
ooijoyce295@gmail.com

Alan W. C. Tan
Faculty of Engineering and Technology
Multimedia University
Melaka, Malaysia
wctan@mmu.edu.my

***Abstract* – This paper focuses on optical music recognition (OMR) system that recognizes the musical symbols on a digitized music sheet and converts them into symbolic music representation. Two main stages are distinguished – pre-processing and symbol analysis. In the pre-processing stage, staves are detected and removed; while in the symbol analysis stage, each musical symbol is recognized and analyzed. The musical semantics are then determined and converted into symbolic music representation stored in text form.**

Keywords – Music recognition, template matching.

I. INTRODUCTION

The optical music recognition (OMR), in general, is the process of converting the music score into a machine-readable format by means of an optical scanner. The digitized image is then analyzed to locate and identify the musical symbols. The research of OMR has been made from way back in 1966: Pruslin's work [1]. His work was then followed up by Prerau, D. S. 1970, who developed an alternative approach by using distinctive bounding box to recognize clefs, rests, accidentals and certain time signatures but not chords [2, 3].

There are various methods used in OMR for symbol recognition. Modayur used morphological methods in his work [4, 5], while Fujinaga used geometrical features and moments including vertical and horizontal projections in his work [6]. Template matching, conversely, is used by Florence Rossant to recognize musical symbols [7], and high level knowledge correction is used in works such as K. C. Ng *et al.* [8] and M. Szwoch [9].

This paper aims to develop a basic musical symbol recognition system without human assistance, and at the same time determine the effectiveness of using projections, morphological methods, and template matching.

II. METHODOLOGY

A. General Overview

In general, there are four stages in music recognition process: stave lines identification, locating musical objects, identification of musical objects, and determination of musical semantics. For the first step, the five stave lines (horizontal lines) are identified and removed. Without the background of the stave lines, the isolated musical symbols are then located. After locating them, these symbols are identified – whether they are notes, rests, clefs, accidentals etc. The final step is to identify the musical semantics based on the symbols found in the previous step.

As illustrated in Figure 1, the processing flow can be divided into three main stages: pre – processing, symbol analysis, and lastly, the output where the results are displayed. It starts with an input, which is an image of the score sheet scanned at a resolution of 300 dpi. The scanned image will be converted into a binary image. Stave lines are detected and removed in the pre – processing stage; whereas in the symbol analysis stage, the notes are analyzed after segmentation. Other symbols such as rests, clefs and accidentals are detected and identified using template matching. Finally, the results will be stored and displayed.

B. Pre – processing

Stave lines contain important information when processing as music is read line by line. By checking the inclination of stave lines, the skew of the image can be corrected.

The scanned image is first converted into a binary image using the iterative threshold selection. After correcting the skew of the image, the stave lines are detected by scanning from left to right. Normally, a music sheet has more than one staff. The image would be scanned from top to bottom, labeling the staff itself so that after five stave lines, it is assumed that a new staff will start. The following operations would also be applied on each of these staff pieces.

After experiencing with different music sheets, it is appropriate to use 80% as threshold to determine the stave lines [10]. From this, the middle line and the stave spacing can be identified, which are useful for later process. After staff identification, the stave lines are removed by replacing the white pixels of each line with black pixels unless there are musical objects detected on either side on the line. If it is more than two pixels away from the top or bottom of that part of the stave line, an assumption will be made that musical objects exist at that region.

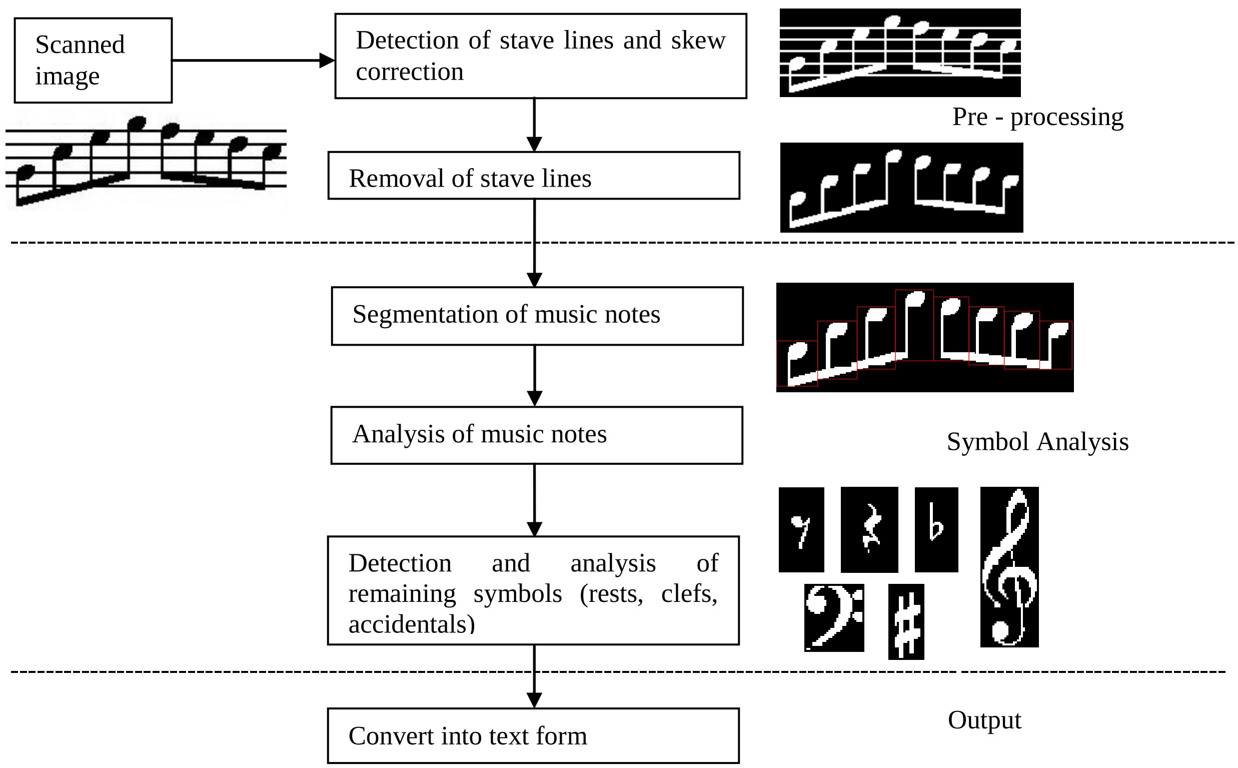


Figure 1. The processing flow

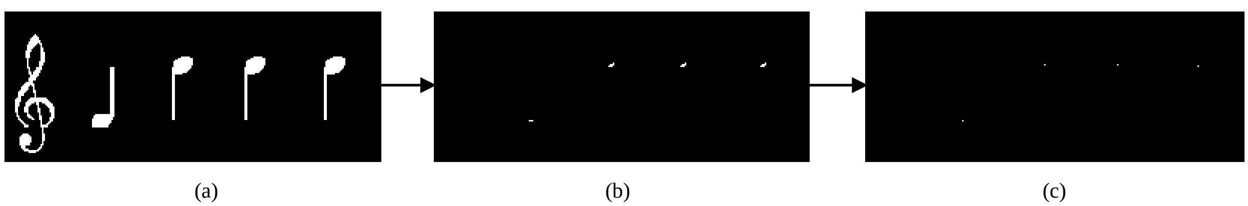


Figure 2. Image (a) is the original image after pre – processing. After eroding image (a), image (b) is produced, and then shrunk to image (c).

C. Symbol Analysis

1) Music Notes Analysis

The music notes are detected by identifying the existence of the note heads. To identify the note heads, morphological techniques are used. Operations such as shrinking and erosion are used to provide shape adjustments on the image. Erosion is used to identify the note head. Erosion is done by scanning the image with the structuring element. The shape of the structuring element used is a disk shaped; while the size of the structuring element is based on the radius of the note head, which is estimated to be half of the stave spacing. The size and shape of the structuring element control the effect of erosion. When the structuring element fits completely inside the object, the probe position is marked. The pixels of each note head are then shrunk to only one pixel coordinate(x, y). Figure 2 shows the morphological process. The position of each note is stored in x and y coordinates accordingly. With these coordinates, the bounding box for each note can be achieved.

The bounding box of each note can be determined by finding the coordinates of the vertical lines and horizontal lines of the bounding box of each note. The vertical lines of the bounding box are achieved by finding the mean of the note head, $x(i)$ and the adjacent note head, $x(i + 1)$. If there is more than one note head in a stem, the x's will be merged first before finding the vertical lines of the bounding box. The horizontal lines of the bounding box, on the other hand, are found by scanning the position of the last white pixel upwards and downwards in between two vertical lines of the bounding box. As each note is bounded in a box, the geometrical features can be extracted through segmentation process. The music notes from the score can be analyzed individually, and with this the pitch of the notes can be identified.

An algorithm is created to find the position of the notes:

$$u = \frac{y - s3}{\frac{d}{2}} \qquad \text{Equation 1}$$

where y refers to the y coordinate of the note head, $s3$ represents the middle line of the staff, and d denotes the stave spacing. Using Equation 1 as stated above with an illustration shown in Figure 3, the position of the note from the middle line of the staff ($s3$) can be determined when the value of u is rounded up.

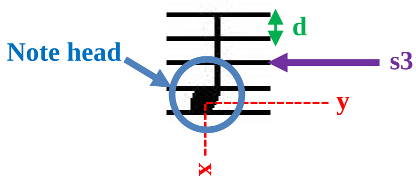


Figure 3. Analysis of the note head

Besides the pitch of notes, the duration of the notes can also be identified. In each bounding box, an analysis of the number of possible flags or beams grouping notes, which are found at the opposite end of the note stem, enables the classification of note duration. Figure 4 illustrates an example of this process.

2) *Template Matching*

Template matching can be done by using a set of models to compute the correlation with the symbol that has to be identified. Templates of rests, clefs, key signature (flats and sharps) and time signature are created and stored respectively to be used for template matching to identify the symbols, with the assumption that the size and shape are fixed as inconsistency of sizes may affect the result of template matching. Using correlation, the symbols on the music score are compared with the symbols in template. If the symbols have perfect matching, the result will reach the maximum score of 1.0. The score decreases as the difference of the number of pixels between the symbols in template and symbols in sheet increase. Different thresholds are used for different objects when template matching is done. For instance, to identify the key signature, the minimum score of 0.85 is used as the threshold while 0.9 is the minimum score to identify the type of rest, and 0.6 is used as the threshold to recognize the type of clef and time signature respectively. The correlation is computed for several coordinates in (x, y). If the score obtained is lesser than the threshold set, the coordinates will not be stored.

III. EXPERIMENTS AND RESULTS

A. *Experiments*

In the first few trials before achieving suitable threshold values, around 20 scanned scores were used to develop this software. Using these training samples, many different algorithms and threshold values were tried to find a suitable recognition rate. The basic strategy for creating this program is based on trial and error. For every trial, if misrecognition is detected, modifications are made to reduce the error rate of symbols analysis. After making the modifications, other samples are tested to check for any side effects of the modification done to ensure that it does not degrade the recognition of other symbols. After that, three samples of music from various publishers are used to check the accuracy of the threshold values. These music samples are shown in the Appendix.

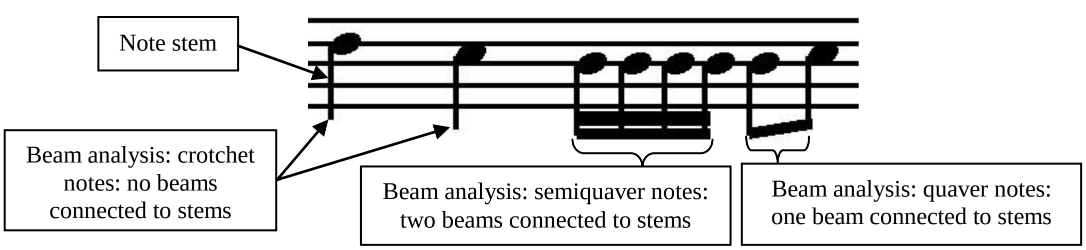


Figure 4. An example of note analysis

TABLE I. THE RESULTS OF THE SAMPLES

Music Samples	*Number of symbols correctly identified*	*Total number of symbols*	*Percentage of recognition rate*
Sample 1 page 1	*131*	*135*	*97.037 %*
Sample 1 page 2	*142*	*150*	*94.667 %*
Sample 2	*171*	*174*	*98.276 %*
Sample 3	*108*	*126*	*85.714 %*
Total	*552*	*585*	*94.359 %*

TABLE II. RECOGNITION RATE BASED ON THE SAMPLES

Symbols	*Notes*	*Rests*	*Clefs*	*Key signature*	*Time signature*
Percentage of recognition rate	*97.283 %*	*88 %*	*100 %*	*77.273 %*	*55 %*

B. *Results*

Tables I and II summarize the recognition results of the tests based on the results of the samples for each main kind of symbols. The recognition rate is computed by finding the total number of stock symbols correctly identified divided by the total number of stock symbols of the music score. A music note is considered correctly identified if and only if both the note duration and pitch of the note are correctly recognized. The typical processing time of the program on an Intel Core 2 Duo 2 GHz is approximately a minute for a page.

C. *Discussion*

Sample 1 is chosen to check the accuracy of recognizing key signature. There are some difficulties in locating the sharps in the key signature, affecting the recognition rate. The 2/4 time signature is well printed. The treble clef is thicker and slightly different compared to the treble clef in Sample 2, but there are no difficulties in recognizing the clefs.

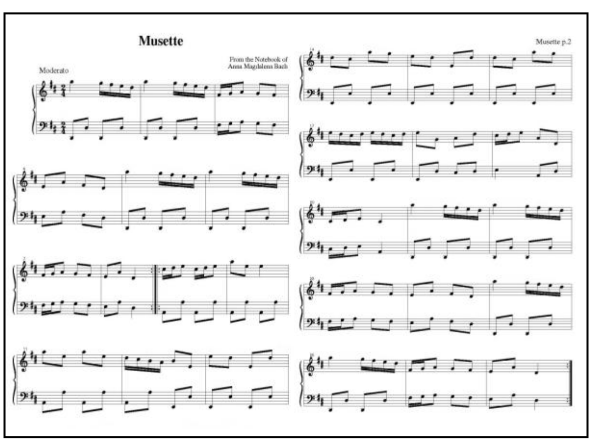
Sample 2 has several rests, making it a good example to use to identify the rests. A few errors were distinguished when detection of rests is done. The time signature is a 2/2 but it is printed as a ‘C’ with a vertical line through it, creating some inaccuracies in the recognition. The treble clef is thinly printed on the stave lines, providing smooth recognition for the clefs.

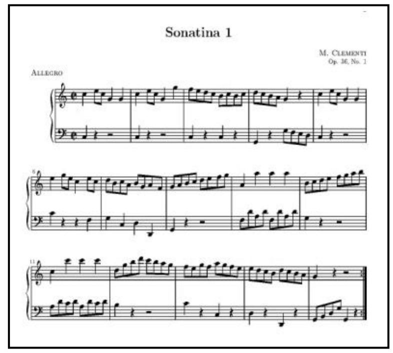
In Sample 3, detection of more than one note head on a stem can be done. As recognizing a single note head on a stem is easier, the recognition rate is affected. Recognition of a different time signature can also be tested as a ‘C’ is used instead of a normal 4/4 time signature.

IV. CONCLUSION

A global recognition system for printed musical scores has been presented in this paper. It starts with staff detection, and followed by symbol analysis. Although the program is undoubtedly still in its infancy, an adequate recognition rate is achieved.

APPENDIX





REFERENCES

[1] Pruslin, D. 1966. Automatic recognition of sheet music. Sc.D. Dissertation, Massachusetts Institute of Technology.

[2] Bainbridge, D., and T. Bell. 2001. The Challenge of Optical Music Recognition. Computers and the Humanities 35(2): 95-121

[3] Prerau, D. S. 1970. Computer pattern recognition of standard engraved music notation. Ph.D. Dissertation, Massachusetts Institute of Technology.

[4] Modayur, B. R., V. Ramesh, R. M. Haralick, and L. G. Shapiro. 1992. MUSER-a prototype musical recognition system using mathematical morphology. Intelligent Systems Laboratory, EE Dept, FT-10: University of Washington.

[5] Modayur, B. R., V. Ramesh, R. M. Haralick, and L. G. Shapiro. 1993. MUSER: a prototype musical recognition system using mathematical morphology. Machine Vision and Applications 6 (2-3): 140-50.

[6] Fujinaga, I., B. Pennycook, and B. Alphonce. 1991. The optical music recognition project. Computers in Music Research 3: 139-42.

[7] Rossant. F. 2001. A global method for music symbol recognition in typeset music sheets. Pattern Recognition Letters 23(10): 1129-1141

[8] Ng, K. C., and R. D. Boyle. 1994. Reconstruction of music scores from primitive Sub-segmentation: School of Computer Studies, University of Leeds

[9] Szwoch, M. Guido: A Musical Score Recognition System. ICDAR 2007: 9th International Conference on Document Analysis and Recognition, Brazil, IEEE Computer Society, pp. 809-813, 2007

[10] Dikici, E. 2006. Musical Note Analysis Using Signal Processing Techniques. Müzik ve Bilim (6)